

# THE RADIUS DISTRIBUTION OF SMALL PLANETS AROUND COOL STARS

TIMOTHY D. MORTON<sup>1</sup>, JONATHAN SWIFT<sup>1</sup>

*Draft version March 14, 2013*

## ABSTRACT

We calculate an empirical, non-parametric estimate of the shape of the radius distribution of small planets with periods less than 90 days using the small yet well-characterized sample of cool ( $T_{\text{eff}} < 4000\text{K}$ ) dwarf stars in the *Kepler* catalog. Using a new technique we call a modified kernel density estimator (MKDE) and carefully correcting for incompleteness, we show that planets with radii  $\sim 1.25 R_{\oplus}$  are the most common planets around these stars. An apparent overabundance of planets with radii  $2\text{--}2.5 R_{\oplus}$  may be evidence for a population of planets with H/He atmospheres. Lastly, the sharp rise in the radius distribution from  $\sim 4 R_{\oplus}$  to  $2 R_{\oplus}$  implies that a large number of planets await discovery around cool dwarfs as the sensitivities of ground-based surveys increase. The radius distribution will continue to be tested with future Kepler results, but the features reported herein are robust features of the current dataset and thus invite theoretical explanation in the context of planetary system formation and evolution around cool stars.

## 1. INTRODUCTION

The discovery of the first exoplanets (Wolszczan & Frail 1992; Mayor & Queloz 1995; Marcy & Butler 1996) has sparked tremendous growth in research and interest in the formation and evolution of planetary systems beyond the Solar System. Not unlike many areas of astronomy, however, the first discoveries are not representative samples; rather, “hot Jupiters” are relatively rare (Wright et al. 2012; Howard et al. 2010) in comparison to the new populations of exoplanets now being revealed by the *Kepler* Mission (Borucki et al. 2011; Batalha et al. 2012; Burke 2013). The most common kinds of planets within *Kepler*’s discovery space of  $R_p \gtrsim 0.5 R_{\oplus}$ , and  $P \lesssim 100\text{d}$  appear to be somewhat larger than Earth but smaller than Neptune,  $1 < R_p < 4 R_{\oplus}$  (Howard et al. 2012; Fressin et al. 2013; Dressing & Charbonneau 2013).

Much of our understanding of planet formation is anchored in decades of research into our own solar system. But now the burgeoning exoplanet population provides us with a new context revealing important insights into planet

formation throughout the Galaxy. For example, the large amount of planetary mass seen close to host stars is evidence that protoplanetary disks may have much higher surface densities than previously thought (Hansen & Murray 2012; Chiang & Laughlin 2012) or that the observed planets migrated from regions further from their host star where more mass was readily available for assembly (Swift et al. 2013).

The multi-transit systems of the *Kepler* sample also offer a wealth of information regarding their formation and evolution (Lissauer et al. 2011; Fabrycky et al. 2012b). The period ratios of planets within a given system show a propensity to lie just outside of first order mean motion resonances (Fabrycky et al. 2012a; Steffen et al. 2013), which may be an imprint of dissipative (Lithwick & Wu 2012; Batygin & Morbidelli 2013) or stochastic mechanisms (Rein 2012) in the formation or evolution of planetary systems. The low inferred mutual inclination of multi-transit systems ( $\sim 1^{\circ}\text{--}3^{\circ}$ ; Fabrycky et al. 2012b; Fang & Margot 2012) together with the relative number of single versus multi-transit systems provides constraints on the number of planets in a given system within *Kepler*’s discovery window, else it may be the first indication of a separate, high-inclination population

tdm@astro.caltech.edu

<sup>1</sup> Department of Astrophysics, California Institute of Technology, MC 249-17, Pasadena, CA 91125

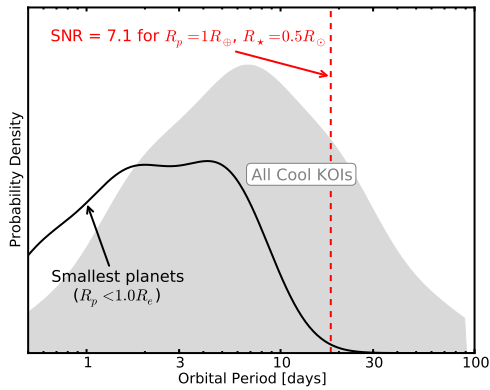


FIG. 1.— Evidence supporting the hypothesis that small planets are incomplete in the Cool KOI sample. The solid black line is the observed (smoothed) distribution of planets smaller than  $1 R_{\oplus}$ ; the grey shaded area is the observed period distribution of all the Cool KOIs. (Neither distribution is corrected for transit probability.) The vertical dashed red line indicates the period at which a  $1 R_{\oplus}$  planet around a  $0.5 R_{\odot}$  star (typical of the Cool KOI sample) would have SNR of 7.1, the nominal detection threshold for KOI identification. The lack of observed small planets at periods longer than 10 days is thus very plausibly due to incompleteness.

of single transit systems (Hansen & Murray 2013; Fang & Margot 2012). Lastly, the mutual gravitational interactions within some multi-planet systems offer an estimation of the planet masses (Lithwick et al. 2012) that then inform planetary compositions and atmospheric evolution scenarios (Rogers et al. 2011; Wu & Lithwick 2012; Lopez et al. 2012).

In this article we focus on yet another important clue regarding the formation of the compact systems revealed by *Kepler*: the distribution of planetary radii. The initial estimates of the planet radius distribution by Howard et al. (2012) showed a dramatic increase in the number of planets at ever smaller size. Citing incompleteness, however, they did not follow this trend in their analysis to planet radii smaller than  $2 R_{\oplus}$ . In an independent study by Youdin (2011), a parametric estimation of the planetary distribution function revealed a deficit of large planets in short period orbits that would support a core accretion then migration formation scenario.

More recent estimates of the planet radius distribution show a preferential size scale in the *Kepler* sample indicated by a flattening

and possible turnover in the log-binned histogram of detected planet candidates somewhere around  $2 R_{\oplus}$  (Fressin et al. 2013; Dressing & Charbonneau 2013; Petigura & Marcy 2013). If true, this would be an important clue toward understanding the key mechanisms that shape the observed population of compact planetary systems that pervade the Galaxy. However, these analyses are constrained by the limitations of coarse histograms; no analysis to date has yet characterized the shape of the exoplanet radius distribution in enough detail to allow meaningful comparison to planet formation and evolution theories.

With this in our sights, we focus on the smallest stars in the *Kepler* Object of Interest (KOI) sample for two reasons: (1) between the spectroscopic studies of the M dwarf sample by Muirhead et al. (2012b) and the photometric re-calibration of the Kepler Input Catalog (KIC) for the coolest stars by Dressing & Charbonneau (2013) the “Cool KOIs” constitute a well characterized sample, and (2) since a transit signal is proportional to the square of the planet to star radius ratio, the Cool KOIs are optimal for probing the planet radius distribution for the smallest planet sizes.

The particular goal of this work is to derive the shape of the planet radius function *properly marginalized over orbital period*. Figure 1 illustrates why this is an issue: the smallest planets are not complete out to the same orbital periods as larger planets; thus careful correction is required in order to achieve this goal. These concepts of incompleteness and period bias were studied in detail early in the history of transit surveys (Pepper et al. 2003; Gould et al. 2003; Gaudi 2005), but have yet to be applied in detail to *Kepler* data.

In §2 we walk through the steps required to properly extract a non-parametric empirical estimate of the true planet radius function given a population detected in a well-characterized transit survey. In §3 we apply these methods to the Cool KOIs to derive the radius distribution for small planets around small stars. We explore the various assumptions that go into this calculation in §5, and conclude in §6.

## 2. FORMALISM

We define the planet radius distribution function  $\phi_r^{P_{\max}}(r)$  such that

$$\int_{r_{\min}}^{r_{\max}} \phi_r^{P_{\max}}(r) dr = \text{NPPS}, P < P_{\max}; \quad (1)$$

that is, a density function with an overall normalization giving the average number of planets per star (NPPS) for planets with period less than  $P_{\max}$  days, for planet radii  $r$  between  $r_{\min}$  and  $r_{\max}$ . The problem of calculating planet occurrence rates from *Kepler* has been quite an industry over the last few years (Youdin 2011; Howard et al. 2012; Dong & Zhu 2012; Swift et al. 2013; Fressin et al. 2013; Petigura & Marcy 2013; Dressing & Charbonneau 2013). However, there has been little quantitative discussion of deriving the detailed shape of the radius function beyond drawing histograms. In the following subsections, we review and refine the general principles of an occurrence calculation and then describe how to follow these principles to construct a non-parametric empirical radius function that obeys the above desired properties.

### 2.1. Occurrence Calculations

In a perfectly idealized survey that is both 100% reliable and 100% complete, the occurrence rate of planets is simply

$$\text{NPPS} = \frac{N_p}{N_{\star}}, \quad (2)$$

where  $N_p$  is the number of detected planets and  $N_{\star}$  is the number of stars surveyed. In practice, however, this must be corrected for both incompleteness and unreliability as follows:

$$\text{NPPS} = \frac{1}{N_{\star}} \sum_{i=1}^{N_p} w_i. \quad (3)$$

Here the sum is over all detections and  $w_i$  is a weighting factor applied individually to account for the various necessary corrections. Generally, these weights can be thought of as

$$w_i = \frac{(1 - \text{FPP}_i)}{\eta_i}, \quad (4)$$

where  $\text{FPP}_i$  is the probability that signal  $i$  is a false positive and  $\eta_i$  is an individualized efficiency factor for the detection of planet  $i$ . In

this work, we do not incorporate the calculations of  $\text{FPP}_i$  in detail, as the *a priori* false positive rate among candidates is low (Morton & Johnson 2011), and ongoing analysis (Morton et al., in prep) according to the false positive-calculating procedure of Morton (2012) indicates the false positive rate in this particular sample is negligibly low.

We thus focus on the detection efficiency  $\eta_i$ , which is defined by the following thought experiment: *If a very large number of planets identical to planet  $i$  were distributed randomly around all the stars in the survey, only a fraction  $\eta_i$  could have been detected.* This can be further factored (following Youdin 2011):

$$\eta_i = \eta_{\text{tr},i} \cdot \eta_{\text{disc},i}, \quad (5)$$

where  $\eta_{\text{tr}}$  is the geometric transit probability, and  $\eta_{\text{disc}}$  is the “discovery efficiency”: the fraction of planets in this thought experiment *with transiting orbital geometries* that could have been detected by the survey. In previous *Kepler* occurrence rate calculations (Howard et al. 2012; Swift et al. 2013; Dressing & Charbonneau 2013), this factor has been defined as

$$\eta_{\text{disc},i} = \frac{N_{\star,i}}{N_{\star}} \quad (6)$$

where  $N_{\star,i}$  is the number of target stars around which planet  $i$  could have been detected. As we show, however, this is not sufficient to properly characterize  $\eta_{\text{disc},i}$ ; more care must be taken.

Central to correctly calculating  $\eta_{\text{disc}}$  is the fact that whatever transit detection pipeline is used, it is incomplete, especially near the detection threshold. This pipeline incompleteness *must* be carefully considered in any occurrence rate calculation. The analysis of Petigura & Marcy (2013) is a model of one way this can be done: simulating planets throughout the radius and period parameter space considered in order to directly measure  $\eta_{\text{disc}}$  as a function of planet radius and period.

A more general conceptual way to attack this problem—and the only one available if one is relying on the results of someone else’s detection pipeline—is to assume that the detection efficiency of any pipeline is a function only of the signal-to-noise ratio (SNR) of the transit signal. This was the approach taken by Fressin et al. (2013), who determined that for the Batalha et al. (2012) Q1-Q6 catalog, the de-

tection efficiency of the *Kepler* pipeline could be modeled by a continuous “SNR ramp” function of the following form:

$$\begin{aligned}\eta_{\text{SNR}}(\text{SNR}) &= 0; & \text{SNR} \leq S_0 \\ &= \text{linear}; & S_0 < \text{SNR} < S_1 \\ &= 1; & \text{SNR} \geq S_1,\end{aligned}\quad (7)$$

where in Batalha et al. (2012)  $S_0 = 6$  and  $S_1 = 16$ . This is notably different from a sharp detection threshold at  $\text{SNR} = 7.1$ , which was used by both Swift et al. (2013) and Dressing & Charbonneau (2013) in their occurrence calculations. The newly released and more uniformly vetted Q1-Q8 KOI catalog, currently hosted at the NASA Exoplanet Archive, is better characterized by a steeper SNR ramp (F. Fressin, priv. comm.), as it is more complete than Batalha et al. (2012); thus, we adopt an SNR ramp where  $S_0 = 6$  and  $S_1 = 12$ .

With this function defined,  $\eta_{\text{disc},i}$  can then be calculated by the following procedure: simulate planet  $i$  around every star in the survey, each of which has both a different radius and different photometric noise properties, to obtain a (normalized) distribution of SNRs  $\phi_{\text{SNR},i}$  for that planet, and then marginalize the detection efficiency over that distribution:

$$\eta_{\text{disc},i} = \int_0^\infty \eta_{\text{SNR}}(s) \cdot \phi_{\text{SNR},i}(s) ds. \quad (8)$$

An extremely important consideration in this procedure of constructing  $\phi_{\text{SNR},i}$  is properly treating orbital period. Remember, the ultimate goal of this analysis is to calculate the distribution of planet radii with period less than  $P_{\text{max}}$ :  $\phi_r^{P_{\text{max}}}$ . Thus, in simulating the population of planet  $i$  clones around other stars, it is important to assign each of these clones an orbital period  $P < P_{\text{max}}$  according to a reasonable estimate of the true planet period distribution  $\phi_P$ .

Distributing the hypothetical planets according to a period distribution is crucial because both SNR and planet occurrence are functions of orbital period. For example, imagine that a survey of 1000 stars detects one  $0.5 R_\oplus$  planet in a 1-day orbit with low SNR (e.g.  $\text{SNR} = 10$ ). An occurrence analysis in the style of Equation 6 might conclude that this planet had a 20% transit probability and would have been detectable around only half the stars in the sur-

vey, thus giving it a weight factor of  $w_i = 10$  and leading to the conclusion that  $0.5 R_\oplus$  planets are rare, only existing around 1% of stars. A slightly more sophisticated analysis might note that  $\eta_{\text{SNR}}(10) = 0.4$ , and give another factor of 2.5 boost, concluding that planets of this size exist around only 2.5% of stars.

However, this conclusion would still be incorrect, since it does not account for the fact that planets of this size may only be detectable at very short periods. What if only a very small fraction of all planets happen to have periods as short as 1 day? *In this case, the supposed rarity of  $0.5 R_\oplus$  planets would be just a misinterpretation of the fact that planets with 1-day orbits are rare.* The small planets that doubtless still do exist at larger orbital periods have not been detected, and no correction has been made to account for this. When constructing  $\phi_{\text{SNR},i}$ , distributing the hypothetical planets according to a reasonable period distribution will avoid this misdiagnosis. This was not done in either Swift et al. (2013) or Dressing & Charbonneau (2013), leading both analyses to underestimate the occurrence rate of small planets around small stars. In §3 we show this also makes a qualitative difference in the interpretation of the planet radius function.

## 2.2. Estimating the Radius Distribution Function

In all the *Kepler* planet occurrence calculations to date, the shape of the radius function has been explored only very coarsely, by calculating the occurrence rate in several different radius bins and either fitting a power law or qualitatively commenting on the shape. Howard et al. (2012) found a good fit to an  $R^{-2}$  power law down to  $2 R_\oplus$ , and declined to comment for smaller planets. On the other hand, Fressin et al. (2013) and Petigura & Marcy (2013) note that the occurrence rate of planets increases towards smaller radius but then appears to flatten out below about  $2.8 R_\oplus$ . Dressing & Charbonneau (2013) claim that the occurrence rate begins to decrease for planets smaller than  $1\text{--}1.4 R_\oplus$ .

Investigating the shape of the radius distribution in more detail requires a non-parametric approach, and also should avoid binning. Here we introduce the concept of a modified kernel density estimator (MKDE) in order to accom-

plish this.

A standard kernel density estimator (KDE) attempts to estimate the true underlying probability distribution of a sample of data points using a function of the following form:

$$\hat{\phi}(x) = \frac{1}{N} \sum_{i=1}^N k(x - x_i; \sigma_i), \quad (9)$$

where  $N$  is the number of data points and  $k(x)$  is a zero-mean, normalized kernel function of arbitrary shape (commonly a Gaussian), with some width  $\sigma_i$ , that most generally can be different for each data point. This creates a smooth distribution out of a discrete data set, with the degree of smoothness controlled by the width parameter. The choice of width has tradeoffs in both directions: if the kernels are too narrow the estimator will be bumpy, but if they are too wide they can wash out real structure in the distribution. Often the width is selected to be the same for all points based on the number of data points, or sometimes a variable-width kernel is used, e.g. the distance to the  $n$ th nearest neighbor. The  $1/N$  normalization factor assures that the integral of this density estimator over the whole parameter space is unity.

In order to use the KDE concept to properly reconstruct the radius function of planets detected in a transit survey, each data point has to be weighted appropriately, leading to a modified KDE, or MKDE:

$$\hat{\phi}_r^{P_{\max}}(r) = \frac{1}{N_{\star}} \sum_{i=1}^{N_p} w_i \cdot k(x - x_i; \sigma_i), \quad (10)$$

where  $w_i = 1/\eta_i$  are the appropriately calculated individual weight factors that renormalize the kernels to correct for missing planets, as discussed in §2.1. The weights ensure that the shape of the radius function responds appropriately to the individual corrections, and the  $N_{\star}$  overall normalization ensures that the integral over all radii will return the NPPS, as desired in Equation 1. A very natural choice for the  $\sigma_i$  in this case, which avoids having to choose an arbitrary smoothing factor, is the uncertainty in each planet’s radius, most of which comes from uncertainty in the radius of the host star. If this does not make for a sufficiently smooth distribution, then the  $\sigma_i$  can be multiplied by

an additional factor to increase the smoothing.

### 3. CALCULATING THE COOL KOI RADIUS FUNCTION

One of the biggest concerns to date about interpreting *Kepler* data is uncertainty about stellar parameters. This applies both because the properties of the transit host stars are unknown (derived planet radius depends directly on the radius of the host star) and because the properties of the stars in the survey parent sample are unknown (i.e. is *Kepler* actually surveying dwarf stars or is the parent sample significantly contaminated by giants or subgiants? (Mann et al. 2012)).

Focusing on *Kepler* candidates around relatively low-mass stars alleviates these concerns. Many of these stars have spectroscopically measured stellar properties (Muirhead et al. 2012a; Mann et al. 2012), and in addition, the properties of the parent sample of target stars has been carefully characterized photometrically by Dressing & Charbonneau (2013). Such an investigation thus is narrower than attempting to use the whole *Kepler* sample, but the assurance of a good understanding of the stellar parameters of both the host stars and the general survey sample more than compensates for this loss of generality. In addition, focusing on these “Cool KOIs” enables detailed study of the radius distribution of Earth-sized and smaller planets.

To construct the planet radius function, we thus select the 113 planet candidates with periods  $< 90$ d identified in the cumulative KOI catalog posted at the NASA Exoplanet Archive that are hosted by stars with  $T_{\text{eff}} < 4000$ K as characterized by Dressing & Charbonneau (2013). To this sample we add the three KOI-961/Kepler-42 planets, which were left out of the Dressing & Charbonneau (2013) sample because its broad-band colors are consistent with classification as either a giant or a dwarf, even though it has been spectroscopically confirmed to be a  $\sim 0.15 M_{\odot}$  dwarf (Muirhead et al. 2012b). For stellar parameters we use the results presented in Dressing & Charbonneau (2013), except for those KOI host stars that have been spectroscopically characterized according to the observations and procedures described in Muirhead et al. (2012b), for which we use the spectroscopic parameters. As this

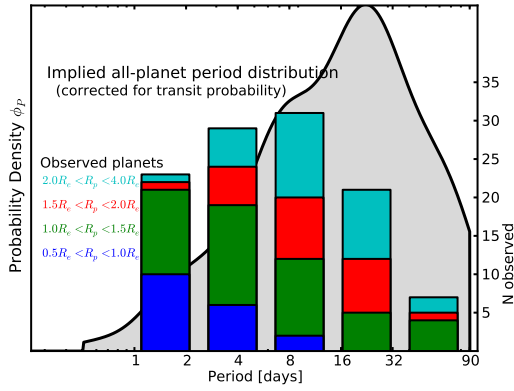


FIG. 2.— The period distribution of planets around *Kepler*’s M dwarfs. The grey shaded region is the implied period distribution of all planets combined, correcting for the effects of transit probability. The bar charts show the observed numbers of planets of different sizes in each period bin. Note the declining fraction of small planets as a function of period—this is most likely an effect of declining detection efficiency for smaller planets on longer-period orbits, and this must be properly accounted for when constructing the planet radius function. The radius function calculation in this paper assumes that all planets are distributed according to the shaded distribution, regardless of planet radius. See §5 for a discussion of this assumption.

spectroscopic method is known to be unreliable for  $T_{\text{eff}} > 3800$  K, we defer to the Dressing & Charbonneau (2013) parameters for stars in this temperature range.

In the following subsections, we describe the steps necessary to calculate  $\hat{\phi}_P^{90}$ , the estimate of the radius function for planets on orbits  $< 90$  d, from this KOI sample. As described in §2, the crucial step toward properly estimating the radius function is calculating the weight factor  $w_i = 1/\eta_i$  for each detection, which includes a transit probability factor and a completeness factor  $\eta_{\text{disc},i}$  (Equation 8). Key to calculating  $\eta_{\text{disc},i}$  is determining the SNR distribution of a hypothetical population of clones of planet  $i$  around all the target stars, or  $\phi_{\text{SNR},i}$ , which in turn requires an assumption of the intrinsic period distribution of planets  $\phi_P$ .

### 3.1. Period distribution

In order to estimate the shape of the true period distribution of planets of all sizes, we make the simplifying assumption that the period distribution of planets is independent of

their radii (see §5 for a discussion regarding this assumption). We thus construct the distribution of  $\log P$  from all the planet candidates in the sample, using an MKDE as described in §2.2. For the weights we use only the inverse transit probabilities, and enforce that the whole distribution is normalized to unity, creating the probability density function for  $\log P$ . For the widths we use  $\sigma = 0.15$  (in  $\log P$ ), to create a smooth distribution. This is the period distribution function  $\phi_P$  that we use in the following subsection, shown as the grey shaded region in Figure 2.

### 3.2. SNR distribution

The SNR of a transit signal is usually defined as follows:

$$\text{SNR} = \frac{\delta}{\sigma} \sqrt{N_{\text{tr}} \cdot N_{\text{pts}}}, \quad (11)$$

where  $\delta$  is the transit depth,  $\sigma$  is the one-point photometric uncertainty,  $N_{\text{tr}}$  is the number of transits observed, and  $N_{\text{pts}}$  is the number of photometric points per transit. It can be shown that, for a fixed planet radius, SNR should scale with host star radius  $R_*$ , orbital period  $P$ , time observed  $T_{\text{obs}}$ , transit duration  $T_{\text{dur}}$ , and  $\sigma$  as follows:

$$\text{SNR} \propto R_*^{-2} P^{-\frac{1}{2}} T_{\text{obs}}^{\frac{1}{2}} T_{\text{dur}}^{\frac{1}{2}} \sigma^{-1}, \quad (12)$$

where  $T_{\text{dur}}$  itself is a function of  $P$ , scaled semi-major axis  $a/R_*$ , and impact parameter  $b$ . The putative SNR of a planet transplanted from its current configuration to a different period, impact parameter, and host star may thus be calculated by scaling the original SNR appropriately.

For each *Kepler* target star, data on the photometric uncertainty  $\sigma$  is available on a quarter-by-quarter basis, quantified by the “combined differential photometric precision” (CDPP) values on 3-hr, 6-hr, and 12-hr time intervals, which in principle should allow for calculation of the SNR for each transit signal. However, the KOI catalogs also provide SNR for each identified planet candidate, and we find that using CDPP values and Equation 11 to calculate SNR does not reliably reproduce the catalog values (it typically underestimates by about 30%, with significant scatter). And since the SNR ramp efficiency characterization (ramp from 0 at SNR = 6 to 1 at SNR

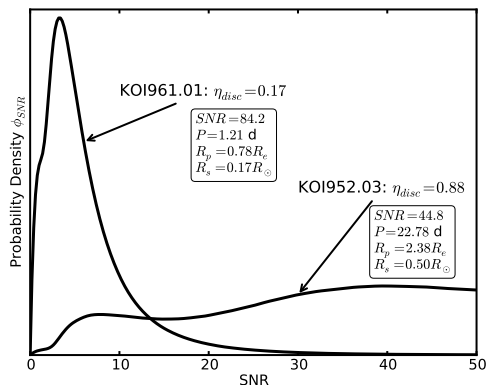


FIG. 3.— Two examples of the SNR distributions resulting from simulating the transit of a given planet around every target star, given randomly assigned periods and impact parameters. The properties of this distribution depend on the properties of the detected system, and the integral of the pipeline detection efficiency function over this distribution gives the “discovery fraction”  $\eta_{\text{disc}}$ . KOI-961.01, a sub-Earth-sized planet in a very short orbit around a very small star (Muirhead et al. 2012b), would have been detectable in only about 1/6 of potential configurations, whereas KOI-952.03, a larger planet around a larger star (Swift et al. 2013), would be detectable in almost any configuration, even though its actual SNR is smaller than that of KOI-961.01.

$= 12$ ) was developed using the catalog SNR values, those are the SNRs we use. We do assume, however, that SNR still scales according to Equation 12.

And so, for each of the planet candidates  $i$  in the Cool KOI sample, we take its SNR as provided by the Exoplanet Archive and construct  $\phi_{\text{SNR},i}$  by simulating the entire population of hypothetical alternative configurations: 10,000 iterations of randomly chosen periods (according to the distribution  $\phi_P$  described in §3.1) and impact parameters (according to a uniform distribution from 0 to 1) around each star in the target population, calculating the appropriate SNR for each instance according to the scalings in Equation 12. As the total number of simulated SNRs is sufficiently large ( $\sim 10^7$ ), the smooth final shape of  $\phi_{\text{SNR},i}$  is defined by interpolating a histogram with bin width  $\Delta \text{SNR} = 1$ . Figure 3 illustrates examples of this distribution for two KOIs.

### 3.3. Radius Distribution

Once  $\phi_{\text{SNR},i}$  is constructed for every planet, we then calculate  $\eta_{\text{disc},i}$  for each planet according to Equation 8, which combined with the transit probability gives the MKDE weight  $w_i$  for each planet, thus building  $\hat{\phi}_r^{90}$  (Equation 10). This function is plotted as the solid black line in Figure 4. To estimate the variance of this density estimator, we perform 1000 different bootstrap resamplings of the true planet dataset and recreate the MKDE for each resampling. The  $1\sigma$  uncertainty region determined by this procedure is illustrated as the grey shaded region in Figure 4, and is conceptually equivalent to a running Poisson error bar. The widths  $\sigma_i$  used to smooth the MKDE are taken to be twice the individual planet radius uncertainties, as this smooths out high-frequency wiggles while retaining broad features. Table 1 presents the data that goes into constructing this distribution.

## 4. RESULTS

The overall normalization of the radius function shown in Figure 4 indicates that there are approximately 1.5 planets per cool star with periods  $< 90\text{d}$ . In addition, there are several notable features of this distribution. The first is the peak between 1 and  $1.5 R_{\oplus}$  and the turnover below, both of which are robust features of the empirical distribution as characterized by the bootstrap uncertainty analysis. If this feature continues to hold as more and more candidates are identified around cool stars, it would point to a dramatic feature of planet formation and evolution:  $\sim 1 R_{\oplus}$  is the most common planet size to survive long-term in short orbits around cool stars. This might be understood by an explanation similar to that provided to explain the origins of the inner Solar System (Goldreich et al. 2004; Chambers 2001): a large number of isolation-mass protoplanets form quickly, and once the gas and planetesimals disk dissipates, a period of dynamical instability follows, at the end of which typically only a few larger planets remain, the rest having been either destroyed (or merged) via collisions or been swallowed by the host star. It is certainly plausible that  $\sim 1 R_{\oplus}$  planets might be the most likely outcome of this process, as this is precisely what has happened with the inner Solar System, with an outcome of two planets about the size of Earth.

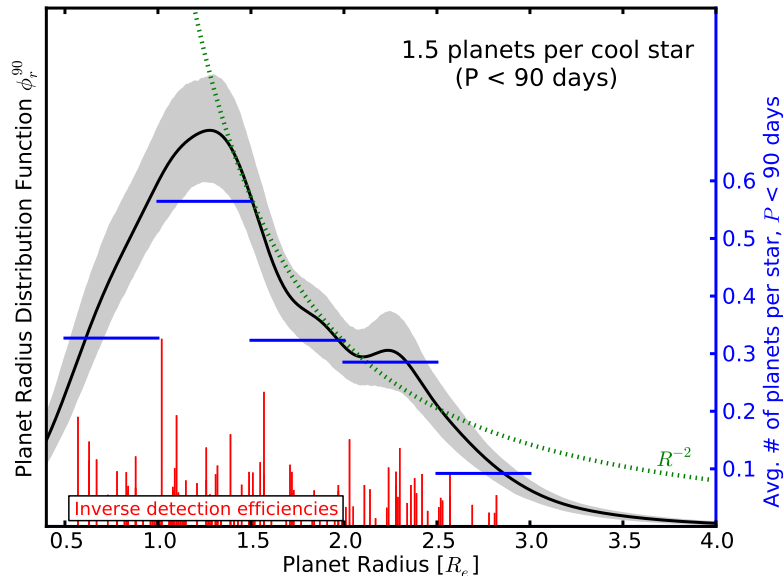


FIG. 4.— The empirical radius distribution of planets orbiting M dwarfs with periods  $< 90$  days (black continuous curve), estimated with a modified kernel density estimator (MKDE; see §3.3), with the bootstrap resampling-derived  $1\sigma$  uncertainty swath shaded grey—essentially a running poisson error bar. The detection efficiency as a function of signal-to-noise ratio has been quantified by an SNR ramp from 0 at SNR = 6 to 1 at SNR = 12. The blue horizontal lines represent the standard “occurrence rate per bin” calculations for this sample. The vertical red lines represent the radii of individual planets in the sample, with their heights being proportional to the weight factors  $w_i$ . The green dotted curve is an  $R^{-2}$  power law, which bears a striking (and uncontrived) resemblance to the shape of this non-parametric radius function between about  $1.25$  and  $2 R_{\oplus}$ ; below about  $1.25 R_{\oplus}$ , the distribution appears to level off, and turn over below  $1 R_{\oplus}$ . Between about  $2$  and  $2.5 R_{\oplus}$  there appears to be an excess over a smooth distribution; this may be caused by a significant population of planets with H/He atmospheres. There is an average of 1.5 planets per cool star in orbits  $< 90$  days over this radius range, and there is an average of greater than 0.5 planets per cool star in this period range for radii between 1 and  $1.5 R_{\oplus}$ .

The second notable feature of the distribution is the plateau between about  $2$  and  $2.5 R_{\oplus}$  and the steep decline above. This could plausibly be the effect of atmospheres, with planets massive enough to retain primordial H/He envelopes showing up as an excess population of planets this size compared to what would be expected from extrapolating toward larger radii from  $\sim 1$ – $1.5 R_{\oplus}$ . One prediction of this hypothesis would be that most of the planets smaller than  $2 R_{\oplus}$  are on average more dense than planets between  $2$  and  $3 R_{\oplus}$ , pointing to a smoother underlying mass distribution.

Finally, this distribution indicates that planets larger than  $\sim 3 R_{\oplus}$  are very rare around cool stars, consistent with the findings of RV surveys (Endl et al. 2003; Johnson et al. 2010b,a; Bonfils et al. 2013). There has been one

hot Jupiter identified around a star in this sample (KOI-254b/Kepler-45b Johnson et al. 2012) and another recent discovery of note (Triaud et al. 2013), but such planets are clearly exceptional—the vast majority of close-in planets around cool stars are smaller than  $\sim 3 R_{\oplus}$ . Even Gliese 1214b (Charbonneau et al. 2009), by far the best-studied planet around an M dwarf to date, appears to be an exception to the typical system, as its radius of  $2.7 R_{\oplus}$  falls far down the tail of this distribution. In fact, there are  $\sim 30\times$  more planets smaller than Gl 1214b than there are larger than Gl 1214b—this bodes very well for the future of ground-based surveys, both transit and RV, as they can become more sensitive to smaller planets.

To explore the degree to which the SNR ramp and the period redistribution affect the shape



of the derived radius function, we repeat this analysis using only the Q1-Q6 KOI catalog (Batalha et al. 2012) and out to a period of 50 days, to match with the analysis of Dressing & Charbonneau (2013). We then compare our full analysis to using a strict SNR = 7.1 detection threshold (i.e.  $\eta_{\text{SNR}} = 1$  above SNR = 7.1 and  $\eta_{\text{SNR}} = 0$  below), and also to using an alternative construction of  $\phi_{\text{SNR},i}$  where the period of planet  $i$  is kept fixed. Figure 5 illustrates these alternative estimates of the radius function. Method (2), illustrated by the dashed line, mirrors the analysis of Fressin et al. (2013), who corrected detections for the SNR ramp effect, but not for the period distribution. Method (3), illustrated by the dotted line, mirrors the analysis of Dressing & Charbonneau (2013), who used an SNR = 7.1 threshold and did not correct for orbital period. We show that these corrections make for a nearly 50% increase in the total inferred number of planets/star for  $P < 50$  d and, notably about a factor of two increase in the number of planets smaller than  $1.4 R_{\oplus}$ . From this comparison, we estimate that the true mean number of Earth-sized ( $0.5$ – $1.4 R_{\oplus}$ ) planets in the habitable zones (HZs) of these cool stars is at least twice as high as the lower limit estimated by Dressing & Charbonneau (2013); that is, probably closer to  $\sim 0.30$ , rather than 0.15. Using revised calculations of the HZ, Kopparapu (2013) calculates a rate of  $\sim 0.50$  habitable Earth-sized planets around cool stars; the completeness considerations in this paper should increase that estimate to  $\sim 1$  planet/star.

Figures 4 and 5 also display the results of these calculations in the more traditional format of a histogram of planet occurrence in different radius bins. In order to calculate these histograms, we simply add up all the weights in each of the radius bins. Figure 4 uses linearly spaced bins; Figure 5 uses the same logarithmic bins used by Dressing & Charbonneau (2013) for comparison. There are several qualitative points to note regarding these histograms. The first is that they can be visually deceptive: for example, the dotted histogram in Figure 5 is approximately flat between 1 and  $2 R_{\oplus}$  before decreasing in the  $0.7$ – $1 R_{\oplus}$  bin, even though the smoothed distributions continue rising steadily all the way to  $\sim 1.25 R_{\oplus}$ —a result of logarithmically spaced bins. Secondly, they only pro-

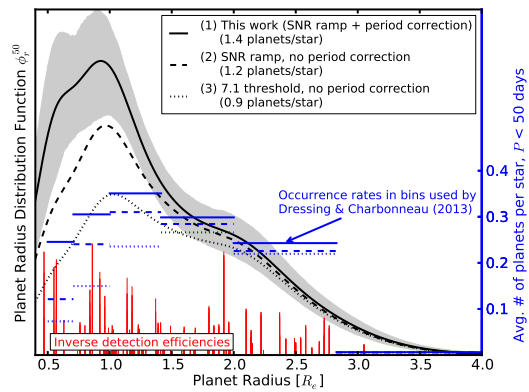


FIG. 5.— The planet radius distribution for  $P < 50$  days, using the Batalha et al. (2012) catalog in order to compare to previous studies, demonstrating the effect of the corrections accounted for in this work. The continuous curves are the non-parametric empirical density estimates, and the horizontal blue lines are the occurrence rates per bin, using the same bins as Dressing & Charbonneau (2013). The vertical red lines represent the radii of individual planets in the sample, with their heights being proportional to the weight factors  $w_i$ . The non-solid linestyles represent different analysis methods. Whereas Method 1 (solid lines) uses the full analysis described in this paper, with detection efficiency described by an SNR ramp (Fressin et al. 2013) and  $\phi_{\text{SNR},i}$  constructed by assigning random periods (§3.2), Methods 2 (dashed) and 3 (dotted) both keep period fixed when constructing  $\phi_{\text{SNR},i}$ , and Method 3 uses an SNR = 7.1 detection threshold rather than the SNR ramp. Method 2 is similar to the occurrence calculation in Fressin et al. (2013) and Method 3 uses the methods employed in Dressing & Charbonneau (2013). The importance of both a well-characterized detection efficiency function and treating period-based incompleteness correctly is clear: incorporating both these considerations significantly changes both the qualitative shape (especially as visualized with histograms) and normalization of the radius function below  $2 R_{\oplus}$ . In particular, the Dressing & Charbonneau (2013) analysis underestimates the occurrence rates of planets between  $0.5$  and  $1.4 R_{\oplus}$  by about a factor of two.

vide a very coarse description of the shape of the radius distribution; that is, there is much more detectable structure than can be captured in a few bins—for example, the plateau in Figure 4 between  $2$  and  $2.5 R_{\oplus}$  is not captured in the histogram illustration, nor is the steep fall-off above  $2.5 R_{\oplus}$ , nor the identification of  $\sim 1.25 R_{\oplus}$  as the location of the low-end turnover.

## 5. EXPLORING ASSUMPTIONS

There are two assumptions that we have made to construct this radius distribution:

TABLE 1  
DATA USED IN RADIUS MKDE

KOI	$R_p$ [ $R_\oplus$ ]	$\sigma_R$	$p_{tr}$	$\eta_{disc}$	$w_i$
KOI961.03 <sup>a</sup>	0.57	0.18	0.051	0.14	140.1
KOI2453.01 <sup>a</sup>	0.63	0.11	0.077	0.12	108.2
KOI2542.01 <sup>a</sup>	0.63	0.08	0.132	0.09	84.2
KOI1422.03 <sup>a</sup>	0.67	0.11	0.051	0.23	85.3
KOI961.02 <sup>a</sup>	0.73	0.20	0.132	0.19	39.9
KOI251.02 <sup>b</sup>	0.78	0.09	0.045	0.55	40.4
KOI961.01 <sup>a</sup>	0.78	0.22	0.068	0.21	70.0
KOI952.05 <sup>a</sup>	0.82	0.08	0.180	0.44	12.6
KOI1843.02 <sup>a</sup>	0.83	0.12	0.044	0.33	68.9
KOI2006.01 <sup>a</sup>	0.84	0.07	0.067	0.33	45.2
KOI2238.01 <sup>a</sup>	0.84	0.13	0.093	0.21	51.2
KOI1146.01 <sup>a</sup>	0.88	0.18	0.034	0.33	89.1
KOI1702.01 <sup>a</sup>	0.88	0.13	0.076	0.27	48.7
KOI250.03 <sup>b</sup>	0.88	0.19	0.059	0.69	24.6
KOI2662.01 <sup>a</sup>	0.92	0.22	0.085	0.33	35.7
KOI2036.02 <sup>a</sup>	0.96	0.16	0.047	0.56	38.0
KOI2306.01 <sup>a</sup>	0.97	0.08	0.241	0.51	8.1
KOI1649.01 <sup>b</sup>	0.98	0.12	0.057	0.52	33.7
⋮	⋮	⋮	⋮	⋮	⋮
KOI255.01 <sup>a</sup>	2.57	0.09	0.016	0.92	67.9
KOI2926.01 <sup>b</sup>	2.57	0.29	0.029	0.90	38.3
KOI248.01 <sup>b</sup>	2.69	0.31	0.041	0.91	26.8
KOI531.01 <sup>b</sup>	2.78	0.40	0.062	0.96	16.8
KOI2156.01 <sup>a</sup>	2.81	0.20	0.068	0.89	16.5
KOI781.01 <sup>a</sup>	2.82	0.10	0.028	0.91	39.2

<sup>a</sup> Planet radius based on spectroscopic stellar parameters from the analysis of Muirhead et al. (2012a)

<sup>b</sup> Spectroscopic stellar characterization not available, or  $T_{eff} > 3800$ , so planet radius based on stellar parameters from Dressing & Charbonneau (2013)

- The period distribution of planets is both independent of planet radius and well-characterized by the current planet detections.
- The detection efficiency of planets in the Batalha et al. (2012) catalog follows an SNR ramp similar to that described in Fressin et al. (2013).

Are these assumptions justified? What are the implications if they are incorrect?

### 5.1. Period Distribution Assumption

Figure 1 illustrates very clearly why the detected population of small planets in the Cool KOI sample is indeed very likely incomplete, showing that where the detected period distribution of the smallest of the Cool KOIs drops off is right around the periods where the known short-period small KOIs would have become undetectable. This is the motivation behind

the period redistribution procedure we use to calculate  $\eta_{disc,i}$  in §3—correcting for the undetectable longer-period small planets. Such a correction is surely needed; however, the nature of this correction as applied in this work—using the implied all-planet period distribution for each planet—merits some discussion.

There are certainly both physical reasons and observational suggestions to believe that the planet period distribution is *not* completely independent of radius. In particular, Howard et al. (2012) finds (shown in their Figure 6) that the fraction of short-period planets that are large ( $4\text{--}8 R_\oplus$ ) is smaller than the fraction of longer-period planets that are large; in other words, the period distribution of larger planets decreases (heading towards shorter periods) sooner than does the distribution of smaller planets ( $2\text{--}4 R_\oplus$ ). Dong & Zhu (2012) present a similar finding. While there is not yet compelling evidence that this same effect has been detected for planets smaller than  $2 R_\oplus$ , simple physical considerations such as increasing stellar insolation (Weiss & Marcy 2013) might reasonably contribute to a dearth of larger planets on short-period orbits. However, as there is no corresponding clear physical explanation for the absence of smaller planets in longer orbits, it is reasonable to assume that they do in fact exist, and that their period distribution might resemble the period distribution of the larger planets that are detected in such orbits.

In addition, if small planets are in general more common than larger planets (as it appears), and small planets are not being detected on longer periods, then approximating the distribution of all planets with just the total observed distribution will naturally *underestimate the total numbers of longer-period planets*. In fact, looking at the all-planet distribution in Figure 2, it is quite reasonable to expect that perhaps the apparent decrease of the distribution function longer than  $\sim 20$  days is actually due to the fact that only planets larger than  $1.5 R_\oplus$  or so (which may very well be a minority of all planets) are being readily detected at these periods. And so, this will cause an overestimate of the true fraction of large planets that are at shorter periods, and an underestimate of the true numbers of small planets that are at longer periods.

The total effect of this assumption will thus

be to systematically shift the simulated “planet clone” distributions towards shorter periods, and thus the SNR distributions  $\phi_{\text{SNR},i}$  toward larger SNRs. This will cause  $\eta_{\text{disc},i}$  values to be slightly overestimated, which will lead to underestimating the weights  $w_i$  and subsequently the normalization of the radius function, especially towards smaller planets, which depend most heavily on this correction. However, this effect is a very small one—we repeated the analysis using a distribution in  $\log P$  that has an exponential cutoff below 10 days and is flat above (that is, without the dip beyond  $\sim 15$  days that appears in Figure 2), and there is negligible difference in the ensuing radius distribution, with the only change being a few percent increase in the planet occurrence in the 1-1.5  $R_{\oplus}$  bin in Figure 4—and notably no change to the 0.5-1  $R_{\oplus}$  bin.

### 5.2. SNR Ramp Assumption

Figure 5 shows that quantifying the detection efficiency of the *Kepler* pipeline as an SNR ramp (Equation 7) rather than a strict threshold cut makes a significant difference in the inferred occurrence rate of planets smaller than 2  $R_{\oplus}$ . In particular, the value of SNR by which the pipeline is assumed to be 100% complete changes the overall normalization of the low end of the radius distribution. While the 6-16 SNR ramp characterization of the efficiency for the Batalha et al. (2012) catalog was introduced and defended by Fressin et al. (2013), and was adjusted in this work (Figure 4) to be relevant to the Q1-Q8 catalog (F. Fressin, priv. comm.), it should be treated as a temporary solution until the detection efficiency of the *Kepler* pipeline can be directly quantified as a function of SNR through injection/recovery simulations. However, the notable features in Figures 4 and 5—the rise down to about 1  $R_{\oplus}$  and the small plateau around 2.5  $R_{\oplus}$ —are robust to the precise details of the ramp.

## 6. CONCLUSIONS

We present a simple non-parametric method of analyzing the empirical shape of the planet radius distribution from a transit survey—the modified kernel density estimator, or MKDE. This estimator is similar to a standard kernel density estimator, except that its overall nor-

malization is constructed to be equal to the total number of planets per star, and that each data point is weighted according to its inverse detection efficiency. We also show that properly computing this efficiency requires two considerations that have not always been applied in previous occurrence rate studies: correcting for the planet period distribution when calculating how many target stars around which a particular planet could have been observed, and considering that the detection efficiency is a rising function of signal-to-noise ratio, and not just a strict cutoff.

Applying this analysis to the 113 planet candidates currently in the cumulative KOI catalog with periods less than 90 days discovered around the cool *Kepler* targets photometrically characterized by Dressing & Charbonneau (2013), we identify several key features of the radius distribution of small planets around small stars which invite theoretical explanation. First, even correcting carefully for incompleteness, the data indicate a flattening or turnover of the distribution around about 1-1.25  $R_{\oplus}$ , suggesting that planets about this size are the most common to survive in short orbits around cool stars. Notably, this feature of the distribution is robust to incompleteness below 1  $R_{\oplus}$ : it appears that planets smaller than Earth are indeed more rare around cool stars than planets around the size of Earth. Secondly, there appears to be a plateau from about 2 to 2.5  $R_{\oplus}$ , where there is an overabundance of planets as compared to what would be predicted from a smooth distribution extrapolating from small to larger planet radii, perhaps an indication of a population of planets with significant H/He atmospheres. And finally, the occurrence pattern of planets around cool stars indicates that there are many planets just beyond the detection threshold of ground-based surveys, as planets larger than Gl 1214b (2.7  $R_{\oplus}$ ) are  $\sim 30\times$  rarer than planets with  $R_p < 2.7 R_{\oplus}$ .

Comparing this non-parametric radius function estimate with the more traditional presentation of planet occurrence rate in different radius bins demonstrates that the bin presentation can be visually misleading, in addition to missing details of the distribution that are accessible in the data. And comparing our results to the occurrence calculations of Dressing

& Charbonneau (2013), we find that there are about a factor of two more planets from 0.5 to 1.4  $R_{\oplus}$  than that analysis determined; this would imply that there are an average of  $\sim 0.30$  habitable-zone Earth-like planets per cool star, rather than the  $\sim 0.15$  estimated by that work. And if this same correction is made to the calculations of Kopparapu (2013), which use updated HZ calculations but the same occurrence formalism as Dressing & Charbonneau (2013), than this number would become closer to  $\sim 1$  planet per star. Habitable-zone, Earth-sized planets abound throughout the Galaxy in numbers even larger than previously estimated.

In addition to demonstrating how to extract empirical distributions from *Kepler* data without relying on arbitrary binning, we call attention to the importance of understanding in detail the detection efficiency of transit search algorithms. Future studies can most directly support analyses such as these—which lie at the very core of the *Kepler* mission—by directly computing the detection efficiency of these algorithms as a function of signal-to-noise ratio, enabling reliable correction for incompleteness near the detection threshold, where many of the most scientifically interesting discoveries will be made. And finally, we emphasize that

this calculation is based on a target sample of only about 3900 cool stars and a KOI search only through Q8 data. Continued expansion of the cool star *Kepler* sample by re-appropriation of target pixels could potentially increase this sample size by a factor of two or more, allowing for greatly strengthened conclusions from the small-planet radius distribution and giving a greater handle on the formation processes of planetary systems around the most numerous stars in the Galaxy. In addition, careful application of these same principles to the entire *Kepler* dataset, as permitted by accurate knowledge of stellar parameters, will continue to uncover important clues to the formation and evolution of all types of planetary systems.

The authors acknowledge John Johnson and the Caltech Exolab for nurturing the intellectual environment in which this work took shape, in particular, conversations with Phil Muirhead, Leslie Rogers, Avi Shporer, Ben Montet, and Jean-Michel Desert. We also thank Peter Goldreich for an enlightening conversation about rocky planet formation, and Dave Charbonneau and Courtney Dressing for helpful comments on an early draft. TDM acknowledges support from NASA NNX11AG85G.

#### REFERENCES

- Batalha, N. M., Rowe, J. F., Bryson, S. T., Barclay, T., Burke, C. J., Caldwell, D. A., Christiansen, J. L., Mullally, F., Thompson, S. E., Brown, T. M., Dupree, A. K., Fabrycky, D. C., Ford, E. B., Fortney, J. J., Gilliland, R. L., Isaacson, H., Latham, D. W., Marcy, G. W., Quinn, S., Ragozzine, D., Shporer, A., Borucki, W. J., Ciardi, D. R., Gautier, III, T. N., Haas, M. R., Jenkins, J. M., Koch, D. G., Lissauer, J. J., Rapin, W., Basri, G. S., Boss, A. P., Buchhave, L. A., Charbonneau, D., Christensen-Dalsgaard, J., Clarke, B. D., Cochran, W. D., Demory, B.-O., Devore, E., Esquerdo, G. A., Everett, M., Fressin, F., Geary, J. C., Girouard, F. R., Gould, A., Hall, J. R., Holman, M. J., Howard, A. W., Howell, S. B., Ibrahim, K. A., Kinemuchi, K., Kjeldsen, H., Klaus, T. C., Li, J., Lucas, P. W., Morris, R. L., Prsa, A., Quintana, E., Sanderfer, D. T., Sasselov, D., Seader, S. E., Smith, J. C., Steffen, J. H., Still, M., Stumpe, M. C., Tarter, J. C., Tenenbaum, P., Torres, G., Twicken, J. D., Uddin, K., Van Cleve, J., Walkowicz, L., & Welsh, W. F. 2012, ArXiv e-prints
- Batygin, K. & Morbidelli, A. 2013, *AJ*, 145, 1
- Bonfils, X., Delfosse, X., Udry, S., Forveille, T., Mayor, M., Perrier, C., Bouchy, F., Gillon, M., Lovis, C., Pepe, F., Queloz, D., Santos, N. C., Ségransan, D., & Berta, J.-L. 2013, *A&A*, 549, A109
- Borucki, W. J., Koch, D. G., Basri, G., Batalha, N., Brown, T. M., Bryson, S. T., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier, III, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Latham, D. W., Lissauer, J. J., Marcy, G. W., Rowe, J., Sasselov, D., Boss, A., Charbonneau, D., Ciardi, D., Doyle, L., Dupree, A. K., Ford, E. B., Fortney, J., Holman, M. J., Seager, S., Steffen, J. H., Tarter, J., Welsh, W. F., Allen, C., Buchhave, L. A., Christiansen, J. L., Clarke, B. D., Désert, J., Endl, M., Fabrycky, D., Fressin, F., Haas, M., Horch, E., Howard, A., Isaacson, H., Kjeldsen, H., Kolodziejczak, J., Kulesa, C., Li, J., Machalek, P., McCarthy, D., MacQueen, P., Meibom, S., Miquel, T., Prsa, A., Quinn, S. N., Quintana, E. V., Ragozzine, D., Sherry, W., Shporer, A., Tenenbaum, P., Torres, G., Twicken, J. D., Van Cleve, J., & Walkowicz, L. 2011, ArXiv e-prints
- Burke, C. 2013, in prep
- Chambers, J. E. 2001, *Icarus*, 152, 205
- Charbonneau, D., Berta, Z. K., Irwin, J., Burke, C. J., Nutzman, P., Buchhave, L. A., Lovis, C., Bonfils, X., Latham, D. W., Udry, S., Murray-Clay, R. A., Holman, M. J., Falco, E. E., Winn, J. N., Queloz, D., Pepe, F., Mayor, M., Delfosse, X., & Forveille, T. 2009, *Nature*, 462, 891

- Chiang, E. & Laughlin, G. 2012, ArXiv e-prints
- Dong, S. & Zhu, Z. 2012, ArXiv e-prints
- Dressing, C. D. & Charbonneau, D. 2013, ArXiv e-prints
- Endl, M., Cochran, W. D., Tull, R. G., & MacQueen, P. J. 2003, *AJ*, 126, 3099
- Fabrycky, D. C., Ford, E. B., Steffen, J. H., Rowe, J. F., Carter, J. A., Moorhead, A. V., Batalha, N. M., Borucki, W. J., Bryson, S., Buchhave, L. A., Christiansen, J. L., Ciardi, D. R., Cochran, W. D., Endl, M., Fanelli, M. N., Fischer, D., Fressin, F., Geary, J., Haas, M. R., Hall, J. R., Holman, M. J., Jenkins, J. M., Koch, D. G., Latham, D. W., Li, J., Lissauer, J. J., Lucas, P., Marcy, G. W., Mazeh, T., McCauliff, S., Quinn, S., Ragozzine, D., Sasselov, D., & Shporer, A. 2012a, *ApJ*, 750, 114
- Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., Rowe, J. F., Agol, E., Barclay, T., Batalha, N., Borucki, W., Ciardi, D. R., Ford, E. B., Geary, J. C., Holman, M. J., Jenkins, J. M., Li, J., Morehead, R. C., Shporer, A., Smith, J. C., Steffen, J. H., & Still, M. 2012b, ArXiv e-prints
- Fang, J. & Margot, J.-L. 2012, *ApJS*, 761, 92
- Fressin, F., Torres, G., Charbonneau, D., Bryson, S. T., Christiansen, J., Dressing, C. D., Jenkins, J. M., Walkowicz, L. M., & Batalha, N. M. 2013, ArXiv e-prints
- Gaudi, B. S. 2005, *ApJ*, 628, L73
- Goldreich, P., Lithwick, Y., & Sari, R. 2004, *ARA&A*, 42, 549
- Gould, A., Pepper, J., & DePoy, D. L. 2003, *ApJS*, 594, 533
- Hansen, B. & Murray, N. 2013, ArXiv e-prints
- Hansen, B. M. S. & Murray, N. 2012, *ApJ*, 751, 158
- Howard, A. W., Marcy, G. W., Bryson, S. T., Jenkins, J. M., Rowe, J. F., Batalha, N. M., Borucki, W. J., Koch, D. G., Dunham, E. W., Gautier, III, T. N., Van Cleve, J., Cochran, W. D., Latham, D. W., Lissauer, J. J., Torres, G., Brown, T. M., Gilliland, R. L., Buchhave, L. A., Caldwell, D. A., Christensen-Dalsgaard, J., Ciardi, D., Fressin, F., Haas, M. R., Howell, S. B., Kjeldsen, H., Seager, S., Rogers, L., Sasselov, D. D., Steffen, J. H., Basri, G. S., Charbonneau, D., Christiansen, J., Clarke, B., Dupree, A., Fabrycky, D. C., Fischer, D. A., Ford, E. B., Fortney, J. J., Tarter, J., Girouard, F. R., Holman, M. J., Johnson, J. A., Klaus, T. C., Machalek, P., Moorhead, A. V., Morehead, R. C., Ragozzine, D., Tenenbaum, P., Twicken, J. D., Quinn, S. N., Isaacson, H., Shporer, A., Lucas, P. W., Walkowicz, L. M., Welsh, W. F., Boss, A., Devore, E., Gould, A., Smith, J. C., Morris, R. L., Prsa, A., Morton, T. D., Still, M., Thompson, S. E., Mullally, F., Endl, M., & MacQueen, P. J. 2012, *ApJS*, 201, 15
- Howard, A. W., Marcy, G. W., Johnson, J. A., Fischer, D. A., Wright, J. T., Isaacson, H., Valenti, J. A., Anderson, J., Lin, D. N. C., & Ida, S. 2010, *Science*, 330, 653
- Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010a, *PASP*, 122, 905
- Johnson, J. A., Gazak, J. Z., Apps, K., Muirhead, P. S., Crepp, J. R., Crossfield, I. J. M., Boyajian, T., von Braun, K., Rojas-Ayala, B., Howard, A. W., Covey, K. R., Schlawin, E., Hamren, K., Morton, T. D., Marcy, G. W., & Lloyd, J. P. 2012, *AJ*, 143, 111
- Johnson, J. A., Howard, A. W., Marcy, G. W., Bowler, B. P., Henry, G. W., Fischer, D. A., Apps, K., Isaacson, H., & Wright, J. T. 2010b, *PASP*, 122, 149
- Kopparapu, R. 2013, arxiv:1303.2649
- Lissauer, J. J., Fabrycky, D. C., Ford, E. B., Borucki, W. J., Fressin, F., Marcy, G. W., Orosz, J. A., Rowe, J. F., Torres, G., Welsh, W. F., Batalha, N. M., Bryson, S. T., Buchhave, L. A., Caldwell, D. A., Carter, J. A., Charbonneau, D., Christiansen, J. L., Cochran, W. D., Desert, J.-M., Dunham, E. W., Fanelli, M. N., Fortney, J. J., Gautier, III, T. N., Geary, J. C., Gilliland, R. L., Haas, M. R., Hall, J. R., Holman, M. J., Koch, D. G., Latham, D. W., Lopez, E., McCauliff, S., Miller, N., Morehead, R. C., Quintana, E. V., Ragozzine, D., Sasselov, D., Short, D. R., & Steffen, J. H. 2011, *Nature*, 470, 53
- Lithwick, Y. & Wu, Y. 2012, *ApJ*, 756, L11
- Lithwick, Y., Xie, J., & Wu, Y. 2012, *ApJS*, 761, 122
- Lopez, E. D., Fortney, J. J., & Miller, N. 2012, *ApJS*, 761, 59
- Mann, A. W., Gaidos, E., Lépine, S., & Hilton, E. J. 2012, *ApJ*, 753, 90
- Marcy, G. W. & Butler, R. P. 1996, *ApJ*, 464, L147+
- Mayor, M. & Queloz, D. 1995, *Nature*, 378, 355
- Morton, T. D. 2012, *ApJ*, 761, 6
- Morton, T. D. & Johnson, J. A. 2011, *ApJ*, 738, 170
- Muirhead, P. S., Hamren, K., Schlawin, E., Rojas-Ayala, B., Covey, K. R., & Lloyd, J. P. 2012a, *ApJ*, 750, L37
- Muirhead, P. S., Johnson, J. A., Apps, K., Carter, J. A., Morton, T. D., Fabrycky, D. C., Pineda, J. S., Bottom, M., Rojas-Ayala, B., Schlawin, E., Hamren, K., Covey, K. R., Crepp, J. R., Stassun, K. G., Pepper, J., Hebb, L., Kirby, E. N., Howard, A. W., Isaacson, H. T., Marcy, G. W., Levitan, D., Diaz-Santos, T., Armus, L., & Lloyd, J. P. 2012b, *ApJ*, 747, 144
- Pepper, J., Gould, A., & Depoy, D. L. 2003, *ApJ*, 53, 213
- Petigura, E. & Marcy, G. 2013, in prep
- Rein, H. 2012, *MNRAS*, 427, L21
- Rogers, L. A., Bodenheimer, P., Lissauer, J. J., & Seager, S. 2011, *ApJS*, 738, 59
- Steffen, J. H., Fabrycky, D. C., Agol, E., Ford, E. B., Morehead, R. C., Cochran, W. D., Lissauer, J. J., Adams, E. R., Borucki, W. J., Bryson, S., Caldwell, D. A., Dupree, A., Jenkins, J. M., Robertson, P., Rowe, J. F., Seader, S., Thompson, S., & Twicken, J. D. 2013, *MNRAS*, 428, 1077
- Swift, J. J., Johnson, J. A., Morton, T. D., Crepp, J. R., Montet, B. T., Fabrycky, D. C., & Muirhead, P. S. 2013, *ApJS*, 764, 105
- Triaud, A. H. M. J., Anderson, D. R., Collier Cameron, A., Doyle, A. P., Fumel, A., Gillon, M., Hellier, C., Jehin, E., Lendl, M., Lovis, C., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smalley, B., Smith, A. M. S., Udry, S., West, R. G., & Wheatley, P. J. 2013, *A&A*, 551, A80
- Weiss, L. & Marcy, G. 2013, in prep
- Wolszczan, A. & Frail, D. A. 1992, *Nature*, 355, 145
- Wright, J. T., Marcy, G. W., Howard, A. W., Johnson, J. A., Morton, T., & Fischer, D. A. 2012, ArXiv e-prints
- Wu, Y. & Lithwick, Y. 2012, ArXiv e-prints
- Youdin, A. N. 2011, *ApJ*, 742, 38